Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2021) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Industry Sciences and Computer Science Innovation

Data Analytics in the Game of Cricket: A Novel Paradigm

Varad Vishwarupe[a]*, Mangesh Bedekar*[b], Prachi M. Joshi[c], Milind Pande[b], Vishal Pawar[b], Prathamesh Shingote[b]

[a]*Amazon Development Centre, India.*
[b]*Maharashtra Institute of Technology WPU, India*
[c]*Vishwakarma Institute of Information Technology, India*

Abstract

Cricket, the second most popular sport in the world, has emerged as one of the most widely played international mainstream sports in the world. With the advent of newer formats such as the Twenty-twenty (T-20) over cricket, the popularity of the game has further soared, spurting out international leagues such as the Indian Premier League (IPL) and the Big Bash League (BBL), over the last one decade. Over the course of more than 450 years since the inception of this sports in south-east England, statistics have played a pivotal role in shaping the structure and format of the game. Of late, there have been considerable advancements in cricket data analytics using IoT-based cricket bat sensors, advanced video analysis using Pitch-Vision, Hawk-Eye and allied technologies. From providing trajectories of the ball, to measuring the speed and impact point of the ball hitting the bat, such stellar breakthroughs have made it possible to fathom and analyze the game in a more refined way. Internet of Things (IoT) and Data Analytics can be the proponents in learning more about the stroke play of players as well as about the myriad facets of the game. In this paper, we have endeavored to discover a novel paradigm of research in cricket analytics known as the timing index, based on a real life IoT-based implementation and case study. The timing index is based on an amalgam of different factors such as bat speed, back lift angle, max. bat speed and impact bat speed. We believe that such a study can help in analyzing a player's training sessions using IoT and data analytics, thereby accurately and holistically analyzing a player's game and going beyond traditional statistical reporting, by ushering in intriguing and insightful information.

Keywords: Cricket Analytics; Cricket Analyst; Artificial Intelligence; Cricket IoT; Cricket Machine Learning; Cricket; Cricket Data Analysis

*Corresponding Authors: varad44@gmail.com, mangesh.bedekar@gmail.com
*Corr. Authors ORCID IDs: www.orcid.org/0000-0002-0372-814X, www.orcid.org/ 0000-0003-4461-9641

1. Introduction

The genesis of cricket lies in South East England, dating back to the 16th century, in Surrey, which has references to this ‘bat and ball’ game. Citing Wikipedia and other references in the upcoming paras, we would first like to introduce the game of cricket in brief. Cricket is a bat-and-ball game played between two teams of eleven players each on a field at the center of which is a 22-yard (20-metre) pitch with a wicket at each end, each comprising two bails balanced on three stumps. The game proceeds when a player on the fielding team, called the bowler, "bowls" (propels) the ball from one end of the pitch towards the wicket at the other end, with an "over" being completed once they have legally done so six times. The batting side has one player at each end of the pitch, with the player at the opposite end of the pitch from the bowler aiming to strike the ball with a bat. The batting side scores runs either when the ball reaches the boundary of the field, or when the two batters swap ends of the pitch, which results in one run. The fielding side's aim is to prevent run-scoring and dismiss each batter (so they are "out", and are said to have "lost their wicket"). Means of dismissal include being bowled, when the bowled ball hits the stumps and dislodges the bails, and by the fielding side either catching a hit ball before it touches the ground, or hitting a wicket with the ball before a batter can cross the crease line in front of the wicket to complete a run. When ten batters have been dismissed, the innings ends and the teams swap roles. The game is adjudicated by two umpires, aided by a third umpire and match referee in international matches. The earliest reference to cricket is in South East England in the mid-16th century. It spread globally with the expansion of the British Empire, with the first international matches in the second half of the 19th century. The game's governing body is the International Cricket Council (ICC), which has over 100 members, twelve of which are full members who play Test matches. The game's rules, the Laws of Cricket, are maintained by Marylebone Cricket Club (MCC) in London. The sport is followed primarily in South Asia, Australia, the United Kingdom, southern Africa and the West Indies. Forms of cricket range from Twenty20, with each team batting for a single innings of 20 overs and the game generally lasting three hours, One Day Cricket comprising of 50 overs to Test matches played over five days. Traditionally cricketers play in all-white kit, but in limited overs cricket they wear club or team colours. In addition to the basic kit, some players wear protective gear to prevent injury caused by the ball, which is a hard, solid spheroid made of compressed leather with a slightly raised sewn seam enclosing a cork core layered with a tight string. [1] Thus, cricket is a highly tactical game in its truest sense and hence has a tremendous scope for applying an analytical viewpoint, especially pertaining to the use of modern technology and data analytics.

2. Related work

Right since the 2000s, the game of cricket has witnessed a significant paradigm shift with the advent of technologies such as the snick-o-meter, hawk-eye etc. There has also been a considerable amount of work on decision making approaches to cricket team selection, predicting the outcome of matches, ranking and rating players on their match-day performance and so on. [2][3][4][5] There has also been some pioneering work on using simulators for simulating cricket matches across different teams and venues, using principal component analysis for prediction, using programming approaches for fantasy league prediction and so on. [6][7][8][9] Statistical approaches such as generating content-based recommendations, personalized recommendations, factor analysis and integer optimization have also yielded some productive results when it comes to the scoring, ranking and prediction of cricket players and matches. [10][11][12][13][14] Batting order is a very important and dynamic part of cricket wherein it is essential to have batsmen playing a certain role viz. anchor, pinch hitter, finisher etc. Top, middle and lower are three parts of the batting order which have designated roles to fulfill so as to win the game, be it while chasing or while batting first. Research has been done on the aforementioned aspects as well, in deciding batting orders based on a plethora of analytical factors. [15][16][17] In the recent past, advances have been made in implementing IoT technologies for gathering vital cricket bat data such as the STR8T, Stance-beam and Bat-sense cricket bat IoT sensors.[18][19] They collect various data points pertaining to the swing of the bat and help the batsmen understand the game better. [20][21][22][23] Cricket analytics has also grown as a field, wherein the data from various matches and tournaments can be analyzed using Python and other languages for churning out important insights related to player performance.[24] While there has been a huge growth in which the game of cricket is played, perceived, analyzed and interpreted, it is still a very nascent field. AI and Machine Learning can especially be used to analyze statistical data in cricket. Combining it with the aforementioned IoT sensors, we believe that a significant revolution can be ushered in the era of cricket analytics.

3. Data Analytics in the Game of Cricket using IoT: A real-life implementation and case study

Cricket bat sensors have revolutionized the way in which cricket data is analyzed and interpreted. The cricket bat sensor is either applied at the top or back of the bat and using the various embedded sensors, it captures the movement of the bat as well as the player during the shot. For the context of this study, we used the aforementioned Stance Beam© cricket bat sensor. Stance Beam is a cricket bat sensor in which a sensor of the size of about 5 cms is attached at the top of the bat. After calculating pitch orientation, make and model of the bat as well as its weight, the sensor is calibrated in a way that it helps to analyze the batsman's shots in real time. For the context of this paper, we tried to simulate real-life playing conditions, albeit at a shorter distance of 18 yards as compared to the conventional 22 yards for the due to the limitations of the sensor and for the sake of getting accurate readings through the bat sensor. The bowling was a mixture of conventional pace bowling and deliveries bowled through a robo-arm, with an average speed of around 120 kmph. The cricket batting net session lasted 15 mins. for the primary author and had 5 overs or 30 balls delivered, out of which the first 6 were used to test and calibrate the sensor. Thereafter we were able to collate and interpret vital data for 24 deliveries, which is equivalent to a conventional net session in cricket clubs. To make the study as real as possible, there were no premeditated shots and everything was impromptu, as is in real cricket matches. To ascertain the effectiveness of the batting session and various batting parameters that were analyzed, it is essential to understand which parameters were taken into account and how were they calculated. As summarized in table 1, the key metrices involved whether the batsman moved forwards (front foot) or backwards (back foot), which is a vital evaluation of a batsman's stroke play in cricket, known as 'footwork'. The next metric involves the different types of shots that were played along with their names. Thereafter, the most important three factors involving the back swing are enunciated which includes the back lift angle, which is the max. horizontal angle achieved during the back lift motion before the start of the downswing. The downswing angle is the angle between the start of the downswing to the point where bat makes an impact with the ball. Lastly, the follow through angle is the motion of the bat after hitting the ball, and is the angle if the bat from the point of impact with the ball to the end of the bat swing. Bat face direction is the point of the bat's face calculated before the start of the swing from the center of the stumps and back lift direction angle is the angle calculated from the center of the stumps up to the stance point before the ball. Lastly, the max speed of the bat swing is taken into account and corroborated with the bat speed at the point of impact of the ball. Since all these quantities are measured in a plethora of units, it is scientifically inadvisable to do any arithmetic operations for calculating the timing index. Hence, we have labelled the timing index on the basis of a holistic overview of these factors and categorized it into 5 categories viz. Early - E, Late - L, Missed - M, Timed - T and Well-Left - W.

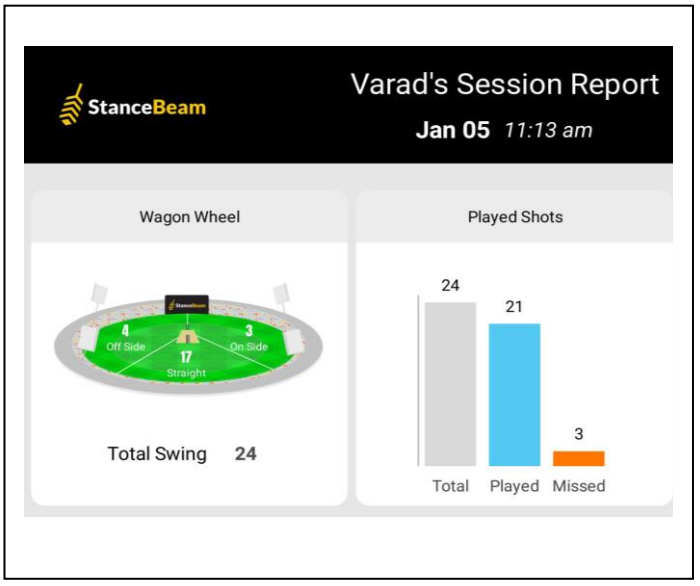

Fig. 1. Session report showing wagon wheel, shots and highlights

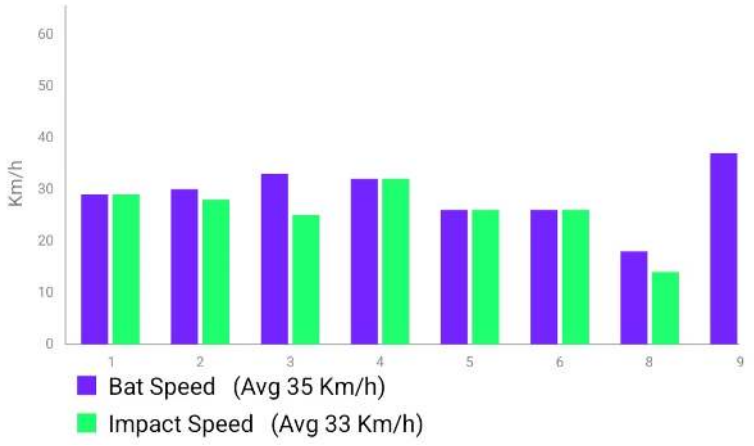

Fig. 2. Bat Speed vs Impact Speed Comparison

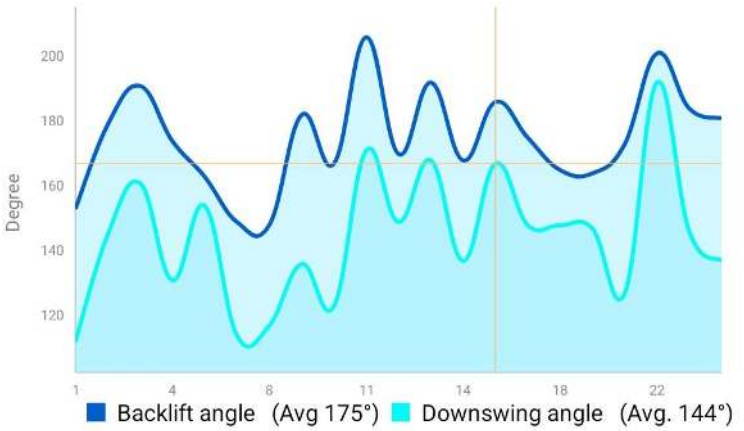

Fig. 3 Backlift Angle vs Downswing Angle

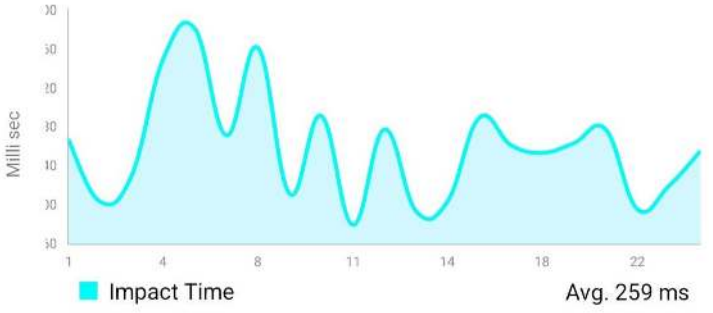

Fig. 4 Impact time of the ball hitting the bat

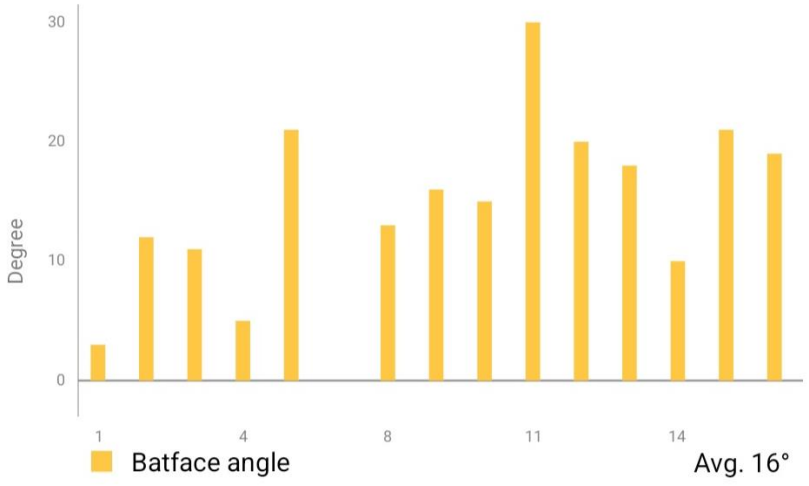

Fig. 5 Bat face angle in degrees

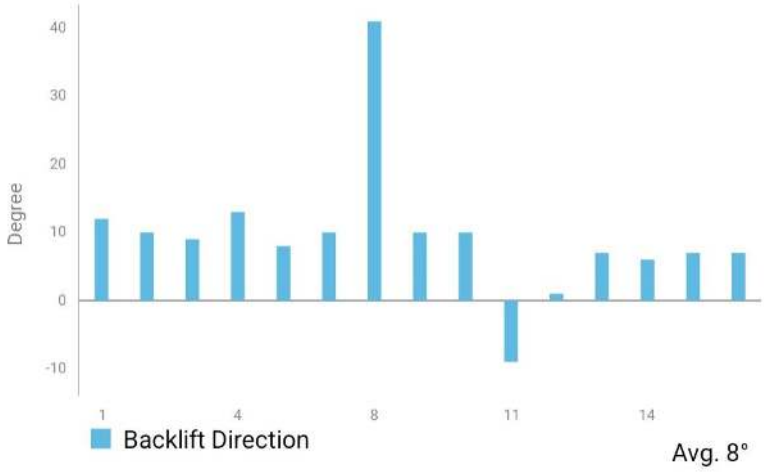

Fig. 6 Bat face direction

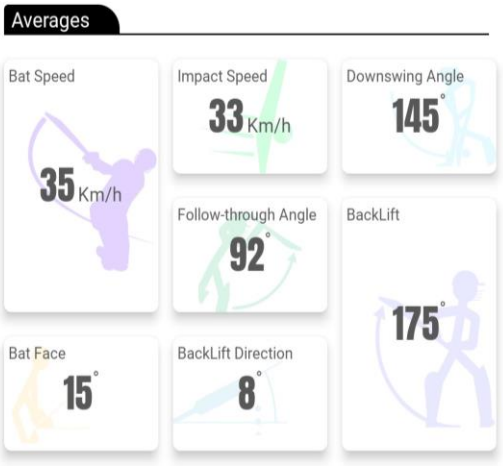

Fig. 7 Session averages and summary

4. Findings, Discussion and Interpretation

Table 1 Shot by shot analysis and timing index categorization

Sr. No.	Shot type (Front/back foot)	Name of the shot played or attempted to play	Back lift Angle, Downswing Angle, Follow Through Angle (degrees)	Bat Face Direction Angle and Back lift Direction Angle (degrees)	Time to Impact after back lift (sec.)	Max. Bat Speed (km/h)	Impact Bat Speed (km/h)	Timing Index (Well Left-W, Early-E, Timed-T, Late-L, Missed-M)
1.	Front	Front foot defence	153,112, 58	03, 12	0.27	29	29	Timed - T
2.	Front	Straight drive	179, 145, 44	12, 10	0.20	30	28	Timed - T
3.	Back	Back foot defence	191, 161, 32	11, 09	0.23	33	24	Late – L
4.	Front	Off drive	129, 174, 74	05, 13	0.35	32	32	Timed – T
5.	Front	Cover drive	163, 154, 15	21, 08	0.38	26	26	Timed - T
6.	Front	Square drive	149, 114, 77	00, 10	0.27	26	24	Timed - T
7.	Back	Upper cut	185, 118, 06	24, -03	0.00	38	00	Well Left - W
8.	Back	Square cut	188, 117, 41	13, 41	0.36	18	06	Early - E
9.	Back	Backward glance	182, 136, 228	16,10	0.21	37	34	Timed - T
10.	Front	Front foot defence	167, 123, 46	15,10	0.29	32	32	Timed - T
11.	Back	Hook shot	206, 171, 239	30, -09	0.18	47	31	Early - T
12.	Front	Front foot defence	170, 149, 87	20, 01	0.28	48	48	Timed - T
13.	Back	Square cut	192, 168, 189	18, 07	0.19	41	29	Early - T
14.	Front	Pull shot	168, 137, 235	10, 06	0.20	42	41	Timed - T
15.	Front	Step out slog	174, 182, 121	10, 09	0.00	52	00	Missed - M
16.	Back	Back foot defence	186, 167, 26	21, 07	0.29	21	11	Late - L
17.	Front	Straight Drive	175, 148, 44	19, 07	0.26	28	24	Timed - T
18.	Front	Flick shot	165, 148, 144	17, 11	0.25	38	37	Timed - T
19.	Front	Cover drive	164, 147, 46	19, 12	0.26	37	19	Early - E
20.	Front	Sweep shot	173, 127, 286	15, 09	0.28	34	33	Timed - T
21.	Back	Back foot drive	182, 123, 126	12, 02	0.00	41	00	Missed - M
22.	Back	Back foot drive	201, 192, 65	20, 02	0.20	57	52	Timed - T
23.	Back	Back foot drive	184, 146, 46	12, 07	0.22	31	19	Early - E
24.	Back	Back foot defence	181, 137, 58	29, 01	0.26	32	30	Timed - T

As seen in Fig. 1 and referring to Table 1, it is evident that a total of 24 shots were played out of which the batsman missed 3 and the ball hit the bat on 21 instances. Furthermore, in cricket, a miss can be categorized as a ‘leave’ known as ‘well left’ when the batsman voluntary chooses to not play a shot and a genuine ‘miss’ when the batsman is beaten by the bowler and is unable to middle the ball at all. As is evident from Figs. 2, 3, 4, 5 and 6, for every shot played, there is a considerable difference in the way the bat sensor records data and this data can be collated and categorized to zero up on the timing index. Starting with the first ball, the batsman attempted to play a total of 11 back foot shots and attempted to play a total of 13 front foot shots. Of these, the batsman missed a couple and voluntarily left one ball. When it comes to the front foot shots, it is observed that the back lift angle is consistently below 180 degrees, and for all back foot shots, its is above 180 degrees, depending on the shot offered. For the correction of back lift trajectories for batsmen, we believe that this is an important finding, as instead of replying solely on visual cues, the batsman can correct his technique based on such qualitative and quantitative findings. Next up, the down swing angle changes and increases as the batsman plays deliveries on the ON side as opposed to the OFF side of the wicket (pitch).

When it comes to the follow through angle, we believe that it is a very subjective and insignificant metric as at that point, energy transfer between the bat and the ball has already taken place and there's no reason to take it into account. Bat face angle and back lift direction angle are more or less the same within a range of 0 to 30 degrees, however they do change and impact the shot when shots such as the Hook, Pull and Sweep are played, which require the batsman to adjust the bat instinctively to allow more 'room' to play the shot by swinging the bat from the off to the on side. The time to impact metric we believe is an important indicator of the batsman's playing ability as for the 62% shots which the batsman timed, keeping other factors same, the batsman took an average of 0.25 seconds to time the ball. It is also observed that shots which are to be played on the backfoot such as the Pull and the Hook require quick reflexes wherein only 0.19 and 0.20 seconds are spent in hitting the ball, from the start of the motion of the back lift. This explains why a large number of batsmen fall prey to the short ball and get caught out by 'mistiming' the ball. Perhaps, the most important and decisive finding of this study involves the max. bat speed and bat speed at impact, wherein there are clear distinctions between different types of timing indices. For every shot that was timed or middled perfectly, there is a difference of less than 5 units, between max. bat speed and bat speed at impact. On the contrary, when the shots were played late, there is marked difference of around 6-10 units in the bat speeds and for the balls wich were played early, the bat speed difference if the highest, upwards of 10 units. This proves that throughout the swing of the bat, there is only a certain range in which the bat speed is the highest, and the batsman should optimally meet the ball at this point, to maximize timing and minimize the difference between max. and impace bat speed. For the ball that was missed, it is seen that although the initial max. bat speed was the highest attributing to the very high backlift angle, the batsman still couldn't meet the ball, due to the miscalculation of the ball's trajectories. Thus, the timing index categorization takes into account the aforementioned parameters and balls that are timed well show that there's a minimal difference between max. and impact bat speed. Moreover, for all the shots that were timed, the batsman shows a steady difference in all the three angular variations of the swing, thereby proving that as much as the speed of the bat swing, direction of the swing for respective shots is equally important. Negative angular ouliers are attributed to the type of shots and creep up only when a certain shot is played, such as the Hook shot or the premeditated Upper cut, which the batsman decided to not play at the last moment. As shown in Fig. 7, eventually, all the session averages are shown which again testify the quantitative nature of batting and how positional, directional and bat swing calculations can help batsmen in perfecting their game and also help us in improving the game's overall understanding.

5. Conclusion and Future Work

Thus, timing index offers a qualitative and holistic overview of a batsman's shot playing abilities and categorization into five pivotal categories which can help in using IoT sensors, which we believe, can be fed to a multi-class label classifier using machine learning algorithms using k-nearest neighbors, decision trees and artificial neural networks, and output these five categories implicitly. For that to happen, however, the normalization of all the quantities mentioned above is a challenge, and can only be done in accordance with the manufacturer's guidelines and at their discretion. All in all, knowing the timing index can especially be important for a batsman when there are no footwork or other changes to the technique, yet, the batsman finds it difficult to middle and time certain deliveries. While we understand that we have honestly attempted to analyze, interpret and partially predict the timing index based on the aforementioned attributes, this study is still a work-in-progress endeavor. Future work in this domain shall involve longer net sessions, sessions involving a spectrum of different players according to the roles such as pinch-hitters, lower order batsmen and opening batsmen. One interesting paradigm for the foreseeable future can involve using cricket gear across a variety of brands, weights and monikers so as to make the study more inclusive and all-encompassing. Another interesting goal of this study can involve varying the pitch length, involving spinners, different types of balls viz. old and new, red, pink and white, and see if such tangible factors can influence the timing index of the batsman. Eventually, batting and cricket being such a exhaustive and gigantic domain, provides a perfect platform for research using IoT, data science, artificial intelligence and machine learning. We believe that it is high time that we move past the tried and tested 'game result prediction' studies involving AI, ML, data science and data mining, and instead focus on using technology as an enabler for improving the qualitative aspects of the game. We believe that, if planned and executed well, there is endless potential in this realm and firmly opine that the confluence of technology and cricket shall redefine the game completely and usher in a new and vibrant era, in the years to come.

References

[1] https://en.wikipedia.org/wiki/Cricket

[2] Ahmed, F., Deb, K. and Jindal, A. (2013). Multi-objective optimization and decision making approaches to cricket team selection. Applied Soft Computing, 13, 402-414.

[3] Akhtar, S., Scarf, P.A. and Rasool, Z. (2015). Rating players in test match cricket. Journal of the Operational Research Society, 66, 684-695.

[4] Allsopp, P.E. and Clarke, S.R. (2004). Rating teams and analysing outcomes in one-day and test cricket. Journal of the Royal Statistical Society Series A, 167, 657-667.

[5] Borooah, V.K. and Mangan, J.E. (2010). The Bradman Class: an exploration of some issues in the evaluation of batsmen for test matches, 187 2006. Journal of Quantitative Analysis in Sports, 6, Article 14.

[6] Preston, I. and Thomas, J.: Batting strategy in limited overs cricket, Statistician, 49(1), p. 95–106 (2000).

[7] Beaudoin, D. and Swartz, T.B. (2003). The best batsmen and bowlers in one-day cricket. South African Statistical Journal, 37, 203-222.

[8] Brettenny, W.J., Friskin, D.G., Gonsalves, J.W. and Sharp, G.D. (2012). A multi-stage integer programming approach to fantasy team selection: A Twenty20 cricket study. South African Journal for Research in Sport, Physical Education and Recreation, 34, 13-28.s

[9] Croucher, J.S. (2000). Player ratings in one-day cricket. In Mathematics and Computers in Sport, G.Cohen and T. Langtry (editors). Bond University, Queensland, Australia, 1-13.

[10] Davis, J. Perera, H., Swartz, T.B. (2015a). A simulator for Twenty20 cricket. Australian and New Zealand Journal of Statistics, 57, 55-71.

[11] Manage, A.B.W. and Scariano, S.M. (2013). An introductory application of principal components to cricket data. Journal of Statistics Education [electronic journal], 21, https://www.amstat.org/publications/jse/v21n3/scariano.pdf.

[12] Davis, J., Perera, H. and Swartz, T.B. (2015b). Player evaluation in Twenty20 cricket. Journal of Sports Analytics, 1, 19-31.

[13] McGinn, E. (2013). The effect of batting during the evening in cricket. Journal of Quantitative Analysis in Sports, 9, 141-150.

[14] V. V. Vishwarupe and P. M. Joshi, "Intellert: a novel approach for content-priority based message filtering," 2016 IEEE Bombay Section Symposium (IBSS), 2016, pp. 1-6, doi: 10.1109/IBSS.2016.7940206.

[15] Staden J.: Comparison of cricketers' bowling and batting performances using graphical displays, Current Science, 96(6), p. 764–766 (2009).

[16] S. Shah, P. Hazarika, J. Hazarika Volume 8, A Study on Performance of Cricket Players using Factor Analysis Approach No. 3, International Journal of Advanced Research in Computer Science, , March – April 2017, ISSN No. 0976-5697

[17] Sharp, G.D., Brettenny, W.J., Gonsalves, J.W., Lourens, M. and Stretch, R.A.: Integer optimization for the selection of a Twenty20 cricket team, Journal of the Operational Research Society, 62, p. 1688-1694 (2011).

[18] Sharma, S.K.: A Factor Analysis Approach in Performance Analysis of T-20 Cricket, Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666 Vol.6, Issue 1 (2013): 69-76(2013).

[19] Bedekar M., Zahoor S., Vishwarupe V. (2016) PeTelCoDS—Personalized Television Content Delivery System: A Leap into the Set-Top Box Revolution. In: Satapathy S., Das S. (eds) Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2, SIST, Springer. https://doi.org/10.1007/978-3-319-30927-9_27.

[20] Lemmer, H.H.: Team selection after a short cricket series, European Journal of Sport Science, DOI: 10.1080/17461391.2011.587895 (2013).

[21] Norman, J.M. and Clarke, S.R.: Optimal batting orders in cricket. Journal of the Operational Research Society (2010) 61, 980-986. doi:10.1057/jors.2009.54 (2010).

[22] Vishwarupe V., Bedekar M., Pande M., Hiwale A. (2018) Intelligent Twitter Spam Detection: A Hybrid Approach. In: Yang XS., Nagar A., Joshi A. (eds) Smart Trends in Systems, Security and Sustainability. Lecture Notes in Networks and Systems, vol 18. Springer, Singapore. https://doi.org/10.1007/978-981-10-6916-1_17

[23] Barr, G.D.I. and Kantor, B.S.: A criterion for comparing and selecting batsmen in limited overs cricket, Journal of the Operational Research Society, 55, p. 1266-1274 (2004)

[24] Saniya Zahoor, Mangesh Bedekar, Vinod Mane, Varad Vishwarupe (2016), Uniqueness in User Behavior While Using the Web. In: Satapathy, S., Bhatt Y., Joshi A., Mishra D. (eds) Proceedings of the International Congress on Information and Communication Technology. Advances in Intelligent Systems and Computing, vol 438. Springer, Singapore. https://doi.org/10.1007/978-981-10-0767-5_24